DOCUMENT RESUME

ED 243 938 TM 840 251

AUTHOR McLarty, Joyce R.
TITLE Dual Objective Achievement Test Selection.
PUB DATE Apr 84
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Tests; *Content Analysis; Criterion Referenced Tests; Educational Objectives; Elementary Education; *Elementary School Curriculum; Norm Referenced Tests; Standardized Tests; *Test Selection; Test Use; *Test Validity
IDENTIFIERS Ceiling Effects; *Riverside Unified School District CA; *Test Curriculum Overlap

ABSTRACT

In 1982-83, Riverside Unified School District (California) selected an achievement test to provide both norm- and criterion-referenced (dual objective) data for its achievement testing program. The test selection process was focused around eight evaluation aspects: prescreening of tests, content matching, instructional evaluation, processing evaluation, psychometric evaluation, input from interested staff and parents, cost analysis, and empirical preparation. Despite considerable effort to blend curricular and technical considerations, only the compromise of functional level testing was able to avoid a direct conflict between the curriculum committee's preference for the Comprehensive Tests of Basic Skills (CTBS) content and format, and the technical committee's concern for its ceiling effects. The CTBS covers 25-71 percent of the curriculum, though it does not assess about 40 percent of the curriculum judged measurable by a paper and pencil test. Although it is possible to wring content-referenced information from a standardized achievement test, the experience in Riverside suggests that the content analysis needs to be thorough and that interpretation of test results relating to the curriculum should be restricted to those specific areas in which sufficient match and coverage have been verified. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED243938

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Joyce McLarty

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

DUAL OBJECTIVE ACHIEVEMENT TEST SELECTION

By
Joyce R. McLarty, Ph.D.
Director
Testing and Evaluation Center
Tennessee State Department of Education

TM 840 251

DUAL OBJECTIVE ACHIEVEMENT TEST SELECTION

With the advent of objectives-referenced testing, school districts' concern for the curricular validity of their testing programs has greatly increased. Unwilling either to sacrifice norm-referenced information or to increase testing time and costs, districts now attempt to select standardized achievement tests which will provide both norm and criterion-referenced data.

In 1982-83, Riverside Unified School District of Riverside, California selected an achievement test to provide both norm and criterion-referenced (dual objective) data for its achievement testing program. Riverside Unified School District annually tests all students in grades 1-7 (about 2,000 students per grade level) on the appropriate grade-level test. No out-of-level testing is done. In addition to providing norm-referenced achievement information, achievement test data is used to select compensatory education students, to classify limited English proficient students, to provide teachers with objectives-referenced information on curriculum mastery, to determine whether students have attained minimum proficiency standards, to assess academic progress, to evaluate curricular programs, to allocate curricular emphasis, and to report to the school board and community.

Riverside Unified School District does all of its own test scoring and processing and many uses of the test information are based on either re-manipulation of the data (for example, rescoring based on selected item subsets for minimum proficiency determination), or matching current test

scores with previous test scores and program participation information. Test selection was therefore charged with retaining the district's capability to meet each of those data needs with a single instrument.

The test selection process was focused around eight evaluation aspects: (1) prescreening of tests; (2) content matching; (3) instructional evaluation; (4) processing evaluation; (5) psychometric evaluation; (6) input from interested staff and parents; (7) cost analysis, and (8) empirical preparation. Instructional, processing and psychometric evaluation and input from interested staff and parents were concurrent evaluations which interacted with one another. Empirical preparation was designed to provide a bridge between the selection and implementation process.

Prescreening of Tests

The Research and Evaluation Department began the test selection process by inviting all companies known to publish achievement tests to submit their tests for consideration by the district. A test review request form and a brief summary of the testing program and selection process were attached to each invitation (see Appendix). The test review request form required the publisher to designate the level and form(s) of the test appropriate for May testing at each grade level, K-8. It also contained a paragraph to be signed by the publisher authorizing the district to conduct a review of the test and its supporting materials and absolving the district from any obligation "beyond normal test security requirements".

While publishers were responding to the test review request, a "core" committee was formed to organize the test selection process. Core commit-

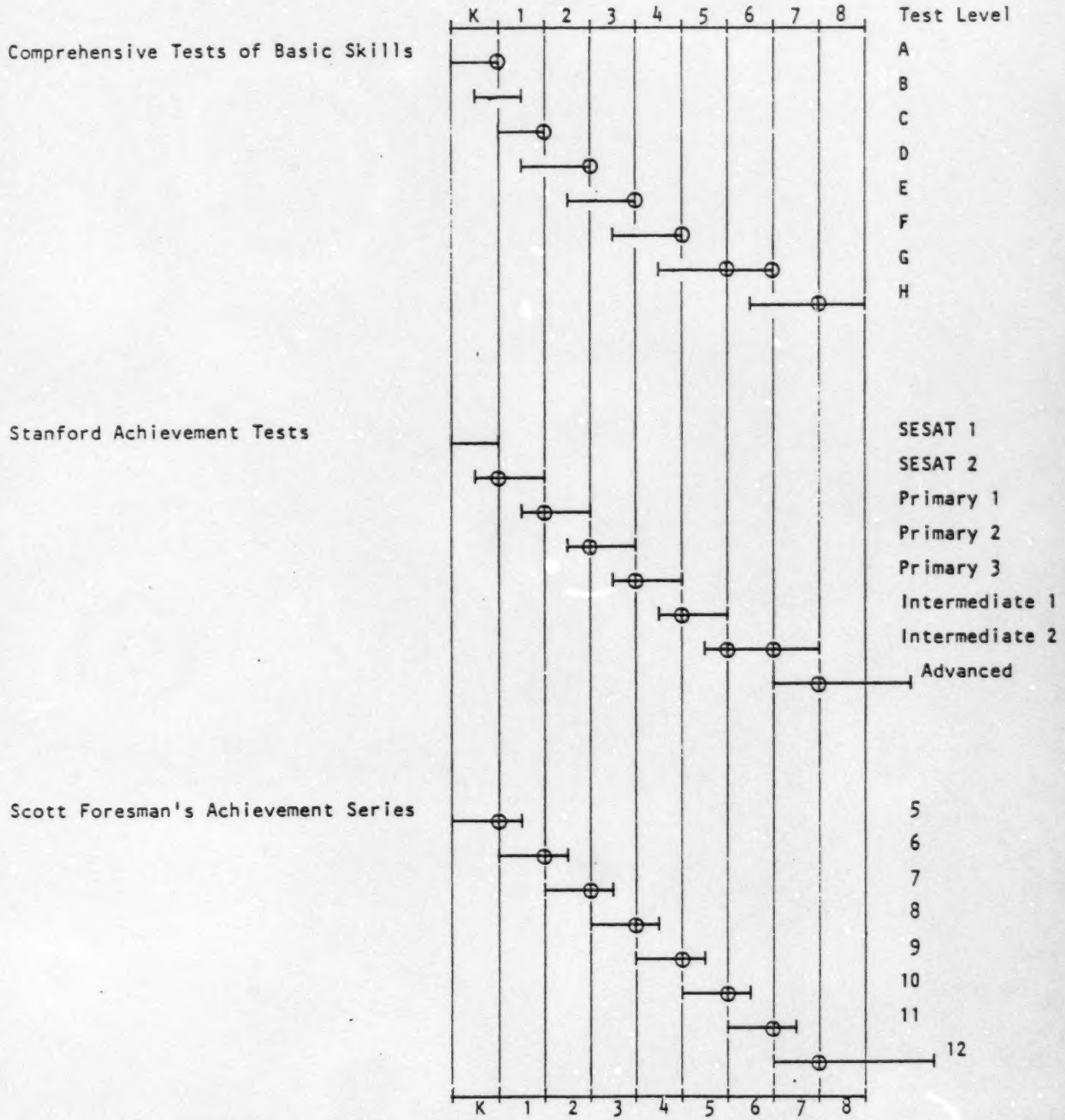
tee members were the three school principals who had most recently chaired textbook selection procedures in the areas of Language Arts, Reading and Mathematics respectively; the Director of Curriculum K-12; the Director of Research and Evaluation; and a Test Specialist from the Research and Evaluation department.

Eleven companies submitted tests for consideration. The test specialist prepared a summary of each test in a standard format. These summaries were reviewed, where possible, by the publishers' representatives. Using these summaries, the core committee prescreened the tests for appropriate general philosophy, content coverage, item selection procedures, general test characteristics, general score characteristics and materials available for review. Eight of the eleven tests were eliminated in this prescreening because of inappropriate content, length or age; or insufficient data for adequate review. The three tests retained by the core committee were the Comprehensive Assessment Program (Scott Foresman); the Comprehensive Test of Basic Skills, Forms U & V, (CTB/McGraw-Hill); and the Stanford Achievement Test, 7th Edition, (The Psychological Corporation).

Table 1 presents a comparison of recommended test levels for May testing for these three tests. The Comprehensive Test of Basic Skills (CTBS) recommended levels are generally at the top of the grade range of each test, allowing no ceiling and about 1 1/2 years of floor. The Scott Foresman Comprehensive Achievement Series (CAP), recommended levels generally allow about a half year of ceiling and a full year of floor. The Stanford Achievement Test (SAT) recommended levels generally allow for about a full year of ceiling and a half year of floor.

TABLE 1

COMPARISON OF TEST LEVELS FOR MAY TESTING



In addition to prescreening the tests, the core committee designed general test selection procedures and established two committees to conduct them. A committee from the Research and Evaluation and Data Processing Departments considered the technical and data processing aspects of the tests, and a committee from the Instruction Division considered the adequacy and appropriateness of content coverage. The Instruction Committee was further divided into three subcommittees: Language Arts, Mathematics and Reading. Members of these sub-committees were selected for their previous involvement in textbook selection, their knowledge of the Riverside Course of Study, and for representation of the schools and grade levels in the district. District office personnel with knowledge of bilingual education and of Chapter 1 and other federally funded programs were also assigned as committee members.

Content Matching

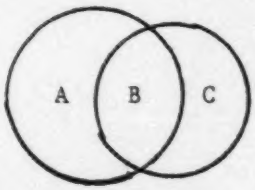
In order to save time, publishers were asked to submit a match of the content of their tests to the Riverside Unified School District Course of Study. As the first step in this matching, the Course of Study was numbered. Each objective, in the course of study was assigned a number from one to the total number of objectives within a subject area for a given grade level. Objectives thought not to be measureable with a written test (for example, oral reading) were not given numbers.

The second step in the matching consisted of having teachers who were familiar with both the Course of Study and the way in which it is taught in Riverside classrooms attempt to match each item in the appropriate test

with one of the objectives in the Course of Study. When a match was found, the item number was written next to the objective in the course of study and the objective number was written next to the item in the test booklet. This process was repeated for every item at every appropriate grade level (kindergarten through seven) in the areas of Reading, Language Arts and Math.

When the item by item matching was complete, the teachers were asked to match the test to the course of study, and to match the course of study to the test (see Figure 1).

Figure 1



Course of Study

Percent of Course of Study covered by the Test = $\frac{B}{A + B}$

Percent of the test matching the Course of Study = $\frac{C}{B + C}$

A form reporting match of the test to the course of study (see Appendix) focused on the degree to which the test covered measurable course of study objectives. For each measurable objective, the items related to that objective were tallied. The percent of measurable objectives covered by one or more items was then calculated. In addition, teachers provided information about the balance of coverage (for example, some objectives

might be covered by only one or two items while others were covered by many items), and about the importance of any unmeasured objectives. If a Course of Study objective was not covered in the test, teachers also judged the seriousness of its omission.

A form matching the Course of Study to the test (see Appendix) looked at the problem from the other side; that is, the degree to which items on the test matched the Course of Study. Items were tallied as matching the Course of Study completely, partially or not at all. Items which matched completely or partially were tallied, totaled and the percent matching reported. In addition, items were categorized as being at a low cognitive level (recall or recognition) or at a higher level (comprehension, application, synthesis or evaluation). Information was reported on the appropriateness of the item format (like or unlike instruction) and on the instructional placement for any unmatched item (whether an item not covered on this test had been covered in a previous grade, was scheduled to be covered in a subsequent grade, or was not covered at all).

Based on these data, tables reporting the percent of Course of Study coverage and the percent of test items matching the Course of Study in Reading, Language and Math, K-7, for each of the three tests were prepared (see Appendix). Overall, the CAP covered 36% of the Course of Study; the CTBS, 51%; and the SAT, 49%. Statistically, coverage by the CAP was significently poorer than that of the other two tests. There were no significant differences in coverage by subject area. In Reading, 43% of the Course of Study was covered; in Language, 44%; and in Math, 49%. Coverage by grade level ranged from 30% at Kindergarten to 58% at third grade, with

average coverage below 50% of the course of study for all but 3 grade levels (3, 4 and 6).

In terms of the match of test items to the Course of Study, the CAP also performed more poorly than the other two tests with 67% of the items matching the Course of Study compared to 76% matching for the SAT and 83% for the CTBS. Matches for the subject areas were also significantly different: 86% for Reading, 65% for Language, and 74% for Math. By grade level, percent matching ranged from 34% at Kindergarten to 91% at grade 2. Statistically, however, Kindergarten was the only significantly different grade level with first grade (the next poorest match) at 67%.

In addition to the Matching Tables, folders were compiled containing the relevant section of the Course of Study and the appropriate test booklet for a given grade level and subject area which had been marked in the matching process. The Course of Study section in the folder contained the original matching item numbers and the test book contained the original matching Course of Study numbers. The folders also contained any notes that the teachers who had completed the content matching wished to have considered by the committee which would make the test selection.

The teachers who conducted the content matching were not otherwise involved in the test selection process. They were selected by the District for their background and knowledge of the curriculum and paid by the publishers for their services. Although they were supervised by the District at no charge, they were otherwise uninvolved in the test selection.

Instructional Evaluation

The Instruction Committee members met four times. Their first meeting was a short after-school orientation. It was followed by two full-day meetings about two weeks apart and a final after school meeting for decision making. After the initial short orientation meeting, the first full day meeting consisted of reviewing each of the tests in detail using the folders prepared by the content matching group. The committee members met in several different formats. Part of the time they met in the subject area sub-committees (Reading, Language, and Math) but they also met in grade level groups so that they could discuss the content of the tests across content areas within specific grade levels. This format allowed for a great deal of discussion and interchange with regard to both content and grade level areas.

At the end of the first full day meeting, the Instruction Committee members met as a whole. At that time they decided to eliminate the Scott Foresman Comprehensive Assessment Program Test from further consideration on the basis of its relatively poor match to the curriculum and the developmental nature of its first few levels. The committee felt that this test simply did not come close enough to meeting the district's needs to warrant the time and energy to review it further.

In the second full day meeting, the Instruction Committee heard presentations from publishers of the two remaining tests, CTB/McGraw Hill and The Psychological Corporation. Each publisher was allowed one hour to present to the committee. Following a lunch break, the committee recon-

vened and rated the two tests using rating forms previously prepared by the core committee. Ratings were on a Likert-type scale. They were in the five general areas of content, item quality, administration, format, and interpretation. The content areas included: match to the course of study, generality over texts, sensitivity to instruction, curriculum coverage, cognitive domain coverage, distribution of coverage, and importance of the content tested. Item quality areas included: curricular process validity, readability level, relevance in interest, clarity, difficulty, bias, currency and examples. The administration areas included: materials design, answer grid design, student identification, time limits, oral items teacher directions and student directions. Format items included: print, spacing, illustrations and use of color. Interpretation included: objectives, scores, and curricular relevance. Each of these areas was further defined in materials prepared for the group (see Appendix).

Ratings of the CTBS and Stanford were generally between 4 and 5 on the five point portion of the scale and between 2 and 3 on the three point portion (see Appendix). The CTBS test was rated slightly higher than the Stanford in almost every area. In the final after-school meeting, the Instruction Committee members reviewed a summary of their ratings and recommended the use of the CTBS test.

Processing Evaluation

Because Riverside school district does all of its own data processing, the compatability of the test booklets and answer sheets to currently used hardware and software was carefully examined. Both tests had the desired

NCS-compatible booklets and answer sheets. Both tests also had primary booklets with more pages than had previously been used. For this reason, it was estimated that primary booklets would take longer to cut and scan than the currently used tests. There was also some question as to whether the test answers could be scored during the scanning process rather than simply recorded and scored on the main frame. However, as both tests were similar in this regard, processing characteristics did not differentiate between these two tests.

Psychometric Evaluation

The Technical Committee, composed of Research Specialists from the Research and Evaluation Department, examined both tests to determine whether either was superior from the standpoint of technical adequacy. Consideration of the tests included examination of the booklets and technical manuals, consideration of norms tables and other printed matter, and meetings with publisher's representatives and technical specialists.

The major technical difference found in examination of these materials was that the CTBS test had been developed as a functional level test which selected items so as to minimize the standard area of measurement in the mid-range of the test. Estimating that the proper range of measurement for a test would be between thirty percent and eighty percent of the items correct, examination of the proper ranges of the CTBS on-level and Stanford tests revealed the probability of severe ceiling effects on the CTBS (see Appendix). The ceiling effects appeared in all subject areas and at all grade levels. Floor effects were found for some grade levels and subject

areas as well, particularly in the upper grade levels. Calculation of the proper range of measurement for the CTBS test one level higher than the on grade test for use with the on grade group revealed that functional level testing of this type would solve the range of measurement problem in most cases. However, seventh grade would require the use of three tests, one below grade level, one on and one above to establish a range of measurement comparible to that found on the Stanford test.

As the Technical Committee reported, the ceiling effect,

> "is exacerbated by the assignment of the median percentile to the group of students who have received the maximum score on a test. A student who misses a single item on a sub-test on an in-level test will often score in the seventieth to eightieth percentile range. A perfect score on an in-level test may result in scores in the eightieth to ninethieth percentile range, especially in the primary grades. Scaled scores are similarly restricted so the net result of a grade level testing program would be to depress scores especially in grades 1, 2, 3 and 7. This would adversely effect both the reliability and validity of scores for higher functioning students and for any averages in which those scores are included".

The Technical Committee recommended that either the CTBS or Stanford test would be acceptable provided that the CTBS, if used, was administered as a functional level test. The Technical Committee noted that the use of CTBS as a functional level test would provide measurement equal and in some cases superior to that on the Stanford but at some cost in content validity. They also noted that the Stanford provided superior measurement at the seventh grade level and in the area of auditory skills for which the CTBS has no comparible section.

Additional Input

Several efforts were made to allow parents, community members and staff not otherwise involved to contribute to the test selection process. District staff explained the existing testing programs and presented the test being considered to the School and Community Advisory Committee. Parents who attended this meeting were encouraged to look over the testing materials and to give comments and suggestions. In general, parental concerns focused on the information provided to parents; particularly in terms of the content of that information, the adequacy of the explanations, and the timing of the score reports. They were especially anxious to have test scores returned to them prior to summer vacation so that if remediation was needed it would be possible to arrange for it during the summer.

Opportunities for staff and other community members to examine the test materials were provided by having a staff member sit with the materials in a meeting room in the district office building for two afternoons after school hours. Teachers in the district were informed that these materials would be available for their review and comment on those afternoons and many visited the office to examine the materials and leave oral or written suggestions. These suggestions were forwarded to the Instructional or Technical Committee as appropriate.

Cost Analysis

Three sets of costs were estimated; one for the Stanford Test, one for the CTBS used as a grade level test and one for the CTBS used as a

functional level test. For each test the costs were identified as initial or recurring costs. Initial costs included the purchase of reusable test booklets and programming for test processing. Recurring costs included the cost for answer sheets, for forms and labels, and for other efforts that could be expected to recur annually.

Cost specific to functional level testing included an added reserve inventory of booklets because of the inability to determine exactly how many students would need each booklet, the use of locator tests and their answer sheets, additional inservice for staff, and an additional allotment for clerical help to get the right test to the right student and back again.

Cost estimates indicated that the CTBS and the Stanford tests would cost approximately the same for on-level testing programs. However, the additional costs associated with functional level testing amounted to an increase in total costs of approximately twenty-two percent. About half of the total cost of grade level testing to be incurred the first year was considered to be recurring cost; however, approximately seventy percent of the total additional cost for functional level testing was considered to be of a recurring nature.

Empirical Preparation

Just as some costs were related to any change of tests while others were related specifically to changing from grade level to functional level testing, the empirical preparation required related in part to any change of tests and in part to the use of functional level testing. The district

had been using the Stanford 6th Edition, so a change to the 7th Edition of that test would be facilitated by equating tables provided by the publisher. Since no equating of the Stanford 6th Edition to the CTBS form was available, the district conducted a small equating study.

Reading and Math scaled scores in grades 1-6, Language scores in grades 5 and 6, and Reading scores in grade 7 were linearly equated. Math scores in grade 7 were equated using an equipercentile method. Results are presented in the Appendix.

On a percentile rank scale, district average scores in Reading and Math for grades 1-7 on the Stanford 6th Edition were compared with projected scores on the CTBS form U and SAT 7th Edition (see Appendix). In Reading, the percentile rank dropped at every grade level for both tests. Drops were larger for the CTBS form U than for the SAT 7th Edition in all grade levels below the seventh.

In Math, percentile increases were projected in grade 7 on the CTBS form U; no change (less than 5 percentile points difference) in grades 3, 4, and 6 on the CTBS form U and grades 5 and 6 on the Stanford 7th Edition, and decreases in grades 1 and 5 on the CTBS form U and grades 1-4 on the Stanford 7th Edition. Large decreases in Math percentiles for both tests were noted at the first-grade level.

Preparation for functional level testing centered on determining which level of the test should be administered to each student. Initial test assignments were planned to utilize the student's previous year's scores to predict whether the student would score outside the "proper range" of the test. Students estimated to score above that range would tentatively be

assigned to the next higher level; those estimated to score below would tentatively be assigned the next lower level. Where no previous test scores were available, or where teachers felt that the tentative assignment was inappropriate, locator tests provided by the publisher could be used.

Since no locator test suitable for use with primary grade level students was available, a 20-item test was constructed to determine whether CTBS/U level C, D, E, or F would be most appropriate.

Additional concerns related to functional level testing related to the adequacy of curricular Math for off-level testing. Although the Course of Study was officially a functional level structure, in practice it was found that Math was often taught at the grade level rather than functional level. Thus, a student who's Reading score indicated the need for a higher level test might not have been instructed in the Math on that test. Similar problems were noted for students who's performance in Reading and Math were substantially different. Students with a home language other than English, for example, might need to take an on-level Math and lower level Reading test.

The logistics of administering and scoring functional level tests assigned by subject area were not feasible for the first year of the new testing program and were, therefore, deferred for future consideration.

Summary

Despite considerable effort to blend curricular and technical considerations, only the compromise of functional level testing was able to avoid a direct conflict between the curriculum committee's preference for

the CTBS content and format and the technical committee's concern for it's ceiling effects. Like many compromises, functional level testing caused additional problems (in this case, increased costs, logistical difficulties, and a possible deterioration of the curriculum match).

The best matched test (CTBS) covered 25% to 71% of the curriculum (less than 60% in 7 of 8 grade levels in Reading, 2 of 8 grade levels in Language and 7 of 8 grade levels in Math). The well-covered curriculum areas tended to be basic skills (e.g., main idea in Reading, capitalization in Language, and multiplication in Math). Less than 90% of the test items matched the curriculum in 3 of 8 grade levels in Reading, 3 of 7 in Language, and 4 of 8 in Math. Thus, while it contained (on the average) less than 10% items which did not match the curriculum, the best matched test did not assess about 40% of the curriculum judged measurable by a pencil and paper test in these areas.

Although it is possible to wring content referenced information from a standardized achievement test, the experience in Riverside suggests that the content analysis needs to be thorough and that interpretation of test results relating to the curriculum should be restricted to those specific areas in which sufficient match and coverage have been verified.

APPENDIX

1. Test Review Request
2. Matching Form: Test to Course of Study
3. Matching Form: Course of Study to Test
4. Match of Scott, Foresman CAP to Course of Study
5. Match of CTBS to Course of Study
6. Match of SAT to Course of Study
7. Definitions for Rating Scale
8. Rating Results
9. Proper Range of Measurement - Reading
10. Proper Range of Measurement - Math
11. Proper Range of Measurement - Language
12. Linear Equating Results
13. Equipercentile Equating Results
14. Percentile Change Projections for CTBS
15. Percentile Change Projections for SAT

Test Review Request

Publisher: _____

Test Name: _____

Authors: _____

Copyright Date: _____

Levels and Forms to Review (May Testing):

Grade	K	1	2	3	4	5	6	7	8
Level									
Form									

Contact Person: _____ Phone: _____

Position: _____

Address: _____

_____

Riverside Unified School District is hereby authorized to conduct a review of the test described above and of any supporting materials (manuals, instruction) for the purpose of determining whether to use it for their achievement testing program. We, the publishers of the named tests, agree to supply Riverside Unified School District with three complete sets of the test and supporting materials. We understand that in accepting these materials, Riverside Unified School District incurs no obligation to the publisher beyond normal test security requirements.

_____

Signature

_____

Date

MATCHING FORM

Match Test to Course of Study:

Course of Study:

Grade: _____ Subject: _____

Test:

Name: _____ Level: _____ Form: _____

Subtest(s): _____

A. Total number of measurable objectives in Course of Study _____

B. List of not-measurable objectives _____

C. For each measurable objective, the number of items which measure it:

(Tally Space)	Number of Items	Tally Total
	10+	
	6-9	
	3-5	
	1-2	
	None*	

D. Total objectives covered on test by 1 or more items _____

E. Percent of measurable Course of Study objectives covered by test (D/B) _____ %

F. *Importance of unmeasured objectives

(Tally Space)	Important	Tally Total
	Low	
	Moderate	
—	Critical	

G. Total items on the subtest(s): _____

H. Number of items matching the Course of Study:

(Tally Space)	Degree of Match	Tally Total
	Complete	
	Partial	
	No Match**	

I. Number of items matching Course of Study completely or partially. _____

J. Percent of test items matching the Course of Study (I/G) _____ %

K. Item process level:

(Tally Space)	Level	Tally Total	Percent
	Recall, Recognition		
	Comprehension, Application, Synthesis, Evaluation		

L. Instructional appropriateness of item format:

(Tally Space)	Format	Tally Total	Percent
	Like Instruction		
	Unlike Instruction		

M.**Instructional placement for unmatched items:

(Tally Space)	Placement	Tally Total
	Previous Grade	
	Subsequent Grade	
	Not Taught	

MATCH OF SCOTT, FORESMAN CAP TO COURSE OF STUDY

Subject	Grade	Course of Study Coverage		Test Items Matching	
		Number of Objectives	Percent Covered	Number of Items	Percent Matching
Reading	K	8	50%	27	67%
	1	30	33%	62	68%
	2	43	37%	95	83%
	3	41	41%	95	82%
	4	38	26%	100	69%
	5	39	28%	100	83%
	6	36	28%	100	83%
	7	87	17%	125	100%
Language	K	1	0%	54	0%
	1	9	0%	27	0%
	2	15	60%	44	89%
	3	17	76%	44	89%
	4	20	45%	57	60%
	5	20	40%	57	58%
	6	20	45%	57	51%
	7	55	22%	65	65%
Math	K	14	0%	15	0%
	1	46	24%	24	37.5%
	2	67	52%	62	100%
	3	75	32%	64	98%
	4	42	57%	75	83%
	5	39	49%	75	71%
	6	40	47.5%	75	73%
	7	67	52%	87	100%

MATCH OF CTBS TO COURSE OF STUDY

Subject	Grade	Course of Study Coverage		Test Items Matching	
		Number Of Objectives	Percent Covered	Number Of Items	Percent Matching
Reading	K	4 (A)	50%	29	57%
		2 (B)	25%	37	24%
	1	20	66%	80	100%
	2	20	49%	83	84%
	3	25	61%	94	99%
	4	19	49%	45	100%
	5	18	46%	120	98%
	6	19	53%	110	100%
	7	24	27%	110	89%
Language	K	0 (A)	0%	1	0%
		0 (B)	0%	1	0%
	1	4	44%	20	80%
	2	10	66%	65	100%
	3	12	71%	70	90%
	4	17	85%	105	81%
	5	14	70%	105	92%
	6	14	70%	105	95%
	7	24	44%	105	76%
Math	K	4 (A)	31%	15	47%
		4 (B)	31%	20	30%
	1	21	46%	45	78%
	2	28	41%	50	96%
	3	49	65%	76	95%
	4	18	43%	85	86%
	5	21	53%	95	98%
	6	21	53%	85	62%
	7	31	46%	85	80%

MATCH OF SAT TO COURSE OF STUDY

Subject	Grade	Course of Study Coverage		Test Items Matching	
		Number Of Objectives	Percent Covered	Number Of Items	Percent Matching
Reading	K	6 (I)	75%	89	100%
		8 (II)	62%	90	74%
	1	18	60%	109	97%
	2	25	58%	121	95%
	3	20	49%	114	93%
	4	15	41%	120	74%
	5	14	36%	120	75%
	6	13	36%	120	70%
	7	19	22%	60	100%
Language	K	Not Tested	Not Tested	Not Tested	Not Tested
	1	2	22%	30	100%
	2	3	2%	30	100%
	3	13	76%	82	84%
	4	13	65%	93	66%
	5	13	65%	103	88%
	6	12	60%	103	67%
	7	13	20%	50	30%
Math	K	9 (I)	64%	40	32%
		5 (II)	36%	50	10%
	1	26	56.5%	79	58%
	2	34	51%	108	74%
	3	38	51%	114	75%
	4	27	64%	118	90%
	5	22	56%	118	78%
	6	29	72.5%	118	74%
	7	48	72%	118	100%

RIVERSIDE UNIFIED SCHOOL DISTRICT
Riverside, California

DEPARTMENT OF RESEARCH AND EVALUATION
December 1, 1982

DEFINITIONS OF CRITERIA FOR RATINGS OF ACHIEVEMENT TESTS

CONTENT

Match to course of study. To what extent do the test items measure objectives included in our course of study for that grade level?

Generality over texts. Will students studying in any one of the district-approved texts be at an advantage (or disadvantage) because of content coverage?

Sensitivity to instruction. Do the items measure skills which can be taught and learned?

Curriculum coverage. Are most of the measureable objectives in the course of study for this grade level covered?

Cognitive domain coverage. Do the items measure a variety of cognitive levels, including applicaton, synthesis and evaluation as well as recall and recognition?

Distribution of coverage. Are the test items well distributed across the course of study objectives for this grade level, or do they all tend to measure a few objectives?

Importance of content tested. Are the most important objectives covered, or only those which are easiest to test?

ITEM QUALITY

Curricular process validity. Do the items mirror the way in which the skills are taught?

Readability level. Are the items written at a level of vocabulary and construction appropriate to the grade?

Relevance/interest. Are the items written so as to be relevant and interesting to the student?

Clarity. Will the student understand clearly what is being asked?

Difficulty. Are the items varied in difficulty. Are some easy items placed first on the subtest?

Bias. Are any items offensive with respect to portrayal of minority groups (sex, ethnic)? Would any items place minority group members at an unfair disadvantage?

Currency. Are any of the items "dated" or in danger of becoming dated rapidly?

Examples. Are sample items clear and appropriate?

ADMINISTRATION

Materials design. Are the materials designed so as to be easy to distribute and collect? Will students be able to locate the appropriate places to start and stop for each subtest?

Answer grid design. Are primary booklets/answer sheets designed so that response spaces are easy to locate and mark, appropriate to the grade level?

Student identification. Are student identification areas on the primary booklets/answer sheets easy to locate and mark? Is the information required clearly specified?

Time limits. Are time limits clearly specified in the directions? Are they sufficient for most students to complete the test?

Oral items. Are directions for administering any items which are to be read to the class clear and complete?

Teacher directions. Is the manual including teacher directions clear and complete?

Student directions. Are directions within the test provided for different types of items, so that students will know clearly what is required?

FORMAT

Print. Is the print size and type clear, easy to read, and appropriate to the grade level?

Spacing. Is the spacing of items on the page appropriate? Are answer spaces, illustrations, etc. clearly associated with the appropriate item (on the same page)?

Illustrations. Are illustrations, graphs and graphics clear and appropriate to the test content?

Color. Is color used appropriately to enhance readability and interest, or does the color (or lack of color) distract or make it difficult to read the test items?

INTERPRETATION

Objectives. Are test objectives clearly and precisely stated? If a student scored poorly on an objective would the teacher know what needed to be retaught?

Scores. Are the different types of scores explained? Are the advantages and limitations of each type noted?

Curricular relevance. Are any materials provided which would assist the teacher in improving student performance for particular objectives or subtests?

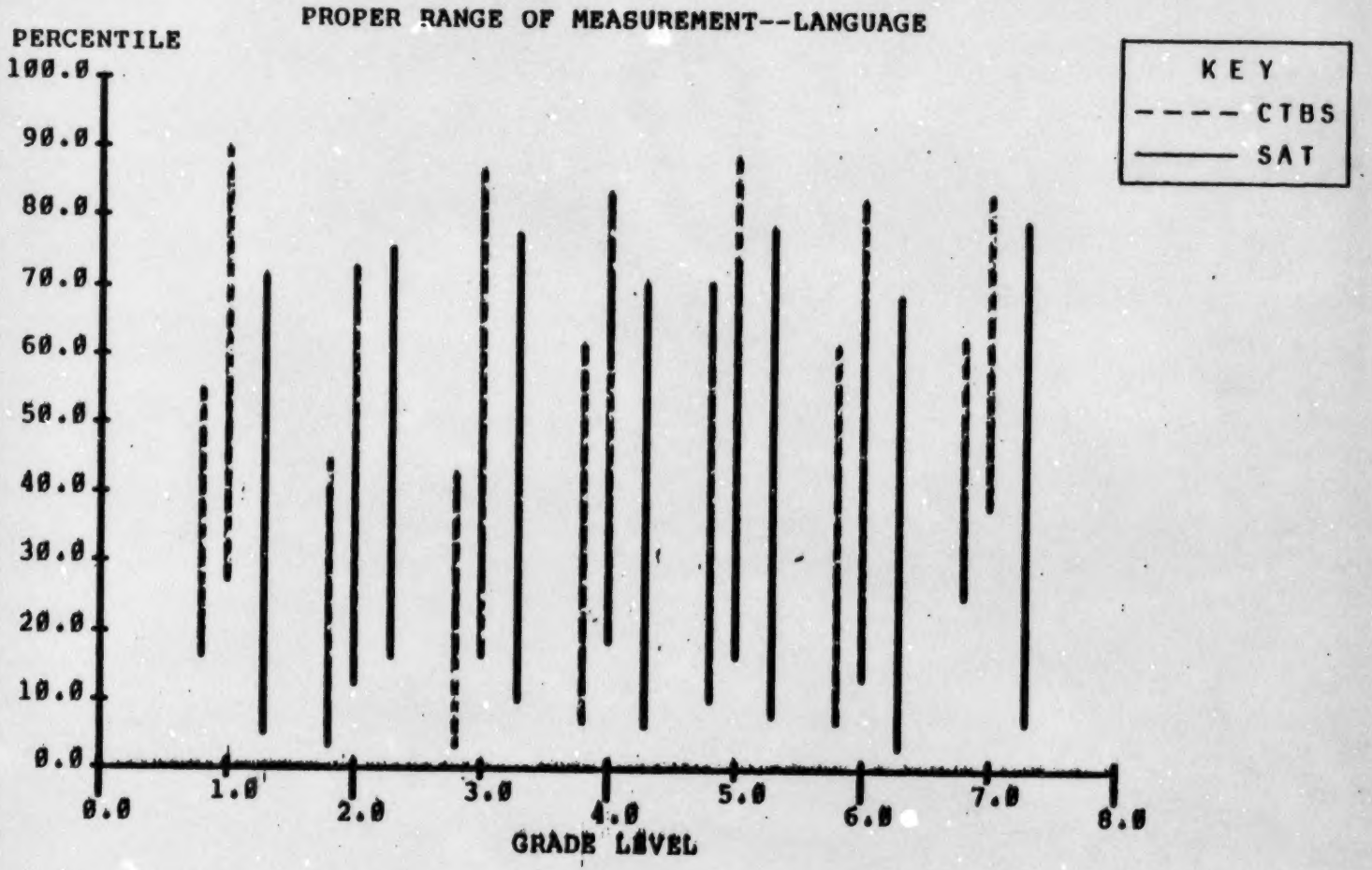
CTBS
SAT

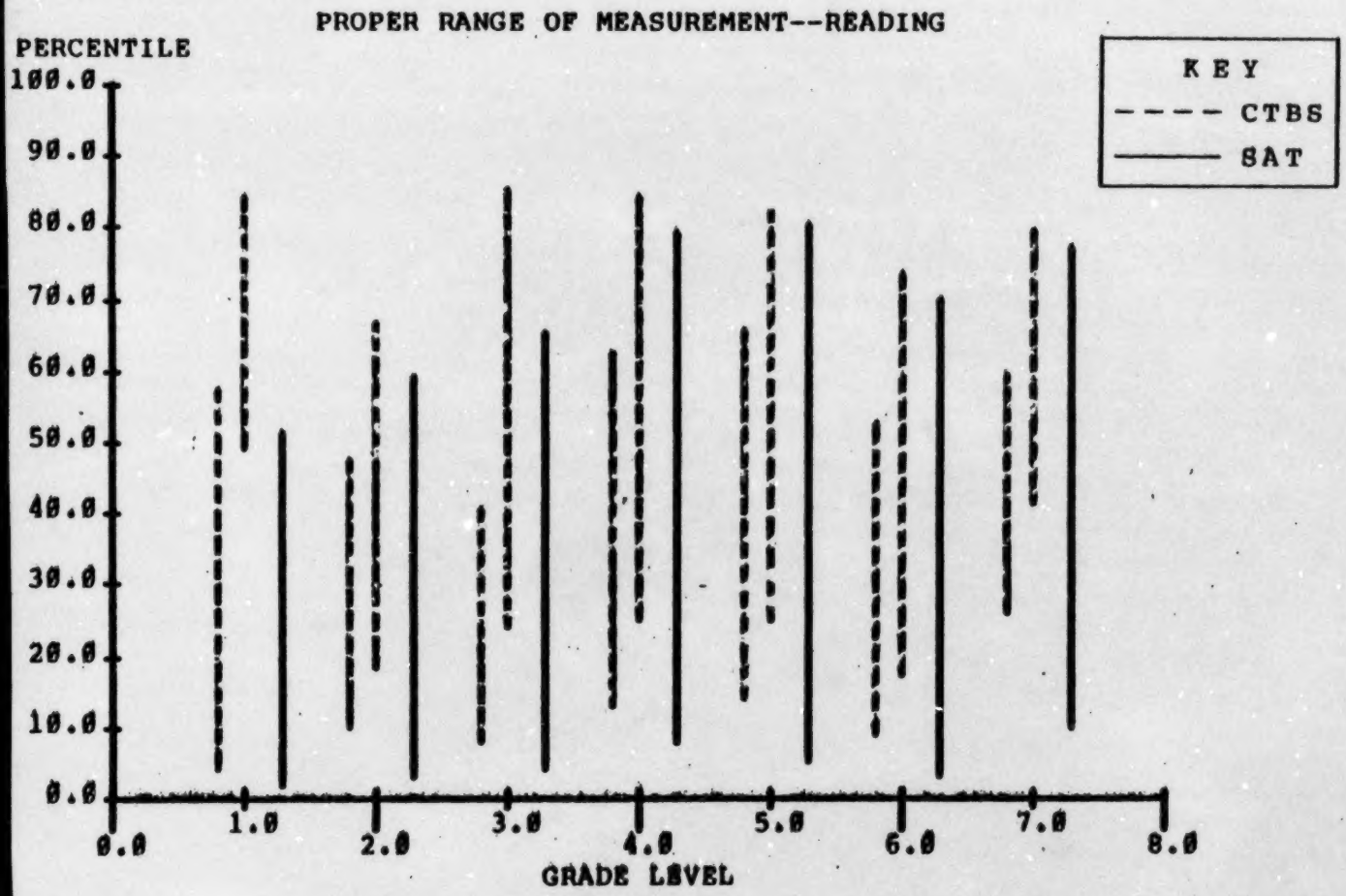
RESULTS OF ACHIEVEMENT TEST RATING

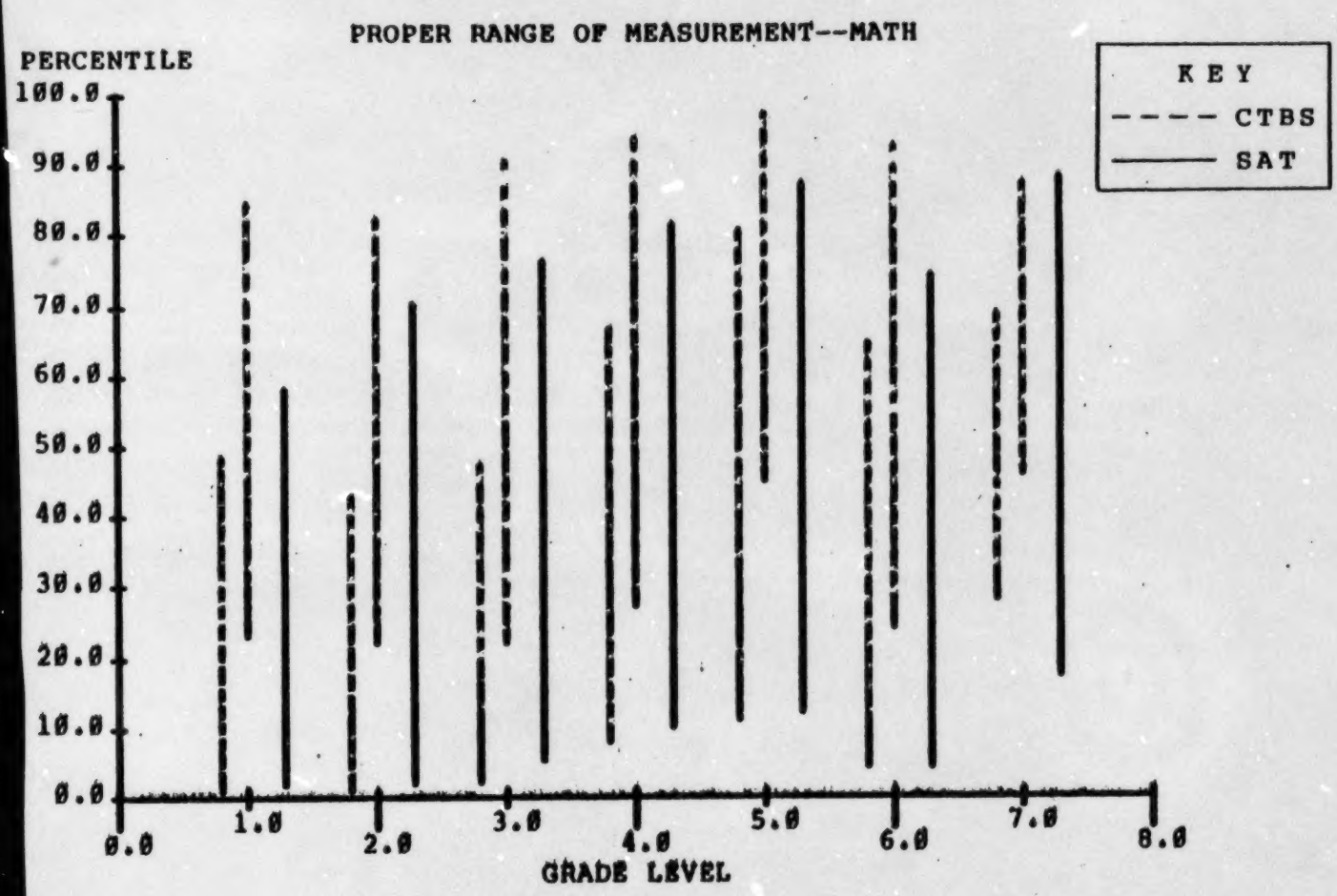
CRITERIA Rate Items 1-9 From 1 (Low) To 5 (High)	1	2	3	4
Content:				
1. How well is the Course of Study covered by the test?				
2. How well are the items on the test covered by the Course of Study?				
3. Is the test well balanced, or are most of the items on a few objectives?				
4. Are both lower and higher level thinking skills tested?				
5. Do the items vary in difficulty?				
6. Are the items directly linked to instruction?				
7. Are the items appropriate to students using different texts?				
8. Is the content tested important (not trivial)?				
9. Do the subtests cover a well-defined content area (not a collection of unrelated skills)?				
CONTENT SUBTOTAL				

Rate Items 10-21 From 1 (Low) To 3 (High)	1	2	3
Item Quality:			
10. Is the readibility level appropriate to the grade?			
11. Are the items relevant and interesting to the grade?			
12. Are the items clearly written?			
13. Are the examples clear and appropriate?			
14. Are the items free of objectional references to race, sex, age, handicap, etc.?			
ITEM QUALITY SUBTOTAL			
Administration:			
15. Will the test be easy to administer?			
16. Will the answer sheets or primary booklets be easy to mark?			
17. Are the timelimits appropriate to the grade level?			
18. Are directions to teachers and students easy to follow?			
ADMINISTRATION SUBTOTAL			

	1	2	3
Format:			
19. Is the print size and spacing appropriate to the grade?			
20. Are the illustrations and use of color appropriate to the grade?			
21. Will the materials withstand handling by students?			
FORMAT SUBTOTAL			
TOTAL			


PROPER RANGE OF MEASUREMENT--LANGUAGE
PERCENTILE
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
KEY
CTBS
SAT
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
GRADE LEVEL


PROPER RANGE OF MEASUREMENT--READING
PERCENTILE
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
KEY
CTBS
SAT
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
GRADE LEVEL


PROPER RANGE OF MEASUREMENT--MATH
PERCENTILE
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
GRADE LEVEL
KEY
CTBS
SAT

Riverside Unified School District
Linear Equating (form y = mx + b) of
Stanford Achievement Test (SAT) and
Comprehensive Tests of Basic Skills (CTBS), form U,
by Grade and Area

Grade	Area	SAT = f(CTBS) m	SAT = f(CTBS) b	CTBS = f(SAT) m	CTBS = f(SAT) b
1	Read	.178	36.4	5.61	-204.2
	Math	.100	77.6	10.01	-776.1
2	Read	.205	19.3	4.88	-94.0
	Math	.142	49.0	7.05	-345.8
3	Read	.248	-4.8	4.04	19.6
	Math	.288	-39.1	3.47	135.6
4	Read	.248	-6.8	4.04	27.6
	Math	.468	-158.6	2.14	338.8
5	Read	.376	-91.8	2.66	243.8
	Lang	.634	-254.9	1.58	402.2
	Math	.769	-358.6	1.30	466.6
6	Read	.366	-83.5	2.73	227.9
	Lang	.690	-299.4	1.45	433.6
	Math	.609	-251.8	1.64	413.8
7	Read	.446	-146.6	2.24	328.9

August 1983

Riverside Unified School District
Total Mathematics, Grade 7
Equipercentile Equating of
Stanford Achievement Test (SAT) and
Comprehensive Tests of Basic Skills (CTBS), form U
via CTBS, form S

SAT	CTBS	SAT	CTBS	SAT	CTBS	SAT	CTBS
145	663	168	705	188	723	203	736
147	666	169	706	189	724	205	737
149	668	170	707	190	725	206	738
151	670	172	708	191	727	208	739
152	673	173	709	192	728	210	740
154	675	175	710	193	729	212	742
156	680	176	711	194	730	213	743
157	685	178	713	195	731	214	744
158	688	180	714	196	732	215	744
160	693	181	716	197	732	216	745
161	697	182	716	198	733	217	746
163	699	183	717	199	734	218	747
164	700	184	719	200	735	220	748
165	701	185	720	201	735	222	750
167	704	187	721	202	735		

August 1983

TABLE 2

CTBS FORM U PERCENTILE PROJECTIONS BASED ON
SAT 6TH EDITION SCORES
May 1983

Average Percentile Rankings

Grade	Reading			Math		
	SAT %ile	CTBS %ile	Diff.	SAT %ile	CTBS %ile	Diff.
1	64	48	-16	74	40	-34
2	60	41	-19	66	69	+ 3
3	58	38	-20	62	61	- 1
4	50	38	-12	58	53	- 5
5	56	39	-17	56	44	-12
6	58	37	-21	58	59	+ 1
7	50	44	- 6	56	69	+13

TABLE 3

SAT 7TH EDITION PERCENTILE PROJECTIONS BASED
ON SAT 6TH EDITION SCORES
May 1983

Average Percentile Rankings

Grade	Reading			Math		
	SAT (6th) %ile	SAT (7th) %ile	Diff.	SAT (6th) %ile	SAT (7th) %ile	Diff.
1	64	53	-11	74	45	-29
2	60	44	-16	66	53	-13
3	58	54	- 4	62	55	- 7
4	50	47	- 3	58	51	- 7
5	56	41	-15	56	55	- 1
6	58	44	-12	58	57	+ 1
7	48	32	-16	56	57	+ 1